Psychological Monographs

General and Applied

No. 359
1953

Moncrieff Smith and Edna A. Wilson

Vol. 67
No. 9

A Model of the Auditory Threshold and Its Application to the Problem of the Multiple Observer

By

Moncrieff Smith
University of Washington

and

Edna A. Wilson
Massachusetts Institute of Technology

Price $1.50

Edited by Herbert S. Conrad
Published by The American Psychological Association, Inc.

Psychological Monographs:
General and Applied

Combining the Applied Psychology Monographs and the Archives of Psychology with the Psychological Monographs

MANUSCRIPTS should be sent to the Editor.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. *Background and bibliographic materials must, in general, be totally excluded* or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

The first page of the manuscript should contain the title of the paper, the author's name, and his institutional connection (or his city of residence). Acknowledgments should be kept brief, and appear as a *footnote* on the first page. No table of contents need be included. For other directions or suggestions on the preparation of manuscripts, see: CONRAD, H. S. Preparation of manuscripts for publication as monographs. *J. Psychol.*, 1948, 26, 447-459.

CORRESPONDENCE CONCERNING BUSINESS MATTERS (such as author's fees, subscriptions and sales, change of address, etc.) should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C.

Vol. 67, No. 9 Whole No. 359, 1953

Psychological Monographs: General and Applied

A Model of the Auditory Threshold and Its Application to the Problem of the Multiple Observer[1]

Moncrieff Smith
University of Washington

and Edna A. Wilson†
Massachusetts Institute of Technology

I. THE THEORY OF THE MULTIPLE OBSERVER

It has often been suggested that two or more independent observers would form a more sensitive detection unit than any one of them alone. This prediction is based on the fact that the sensitivity of an individual varies from moment to moment, so that if one observer is momentarily insensitive, another may detect the signal. In the simplest case, if each of two independent observers has a probability of $\frac{1}{2}$ of hearing a signal, then the probability that both will *fail* to hear it is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$, and the probability that one, the other, or both will hear it is $\frac{3}{4}$.

Schafer (6) has presented the beginnings of a theory of the multiple observer for the case of the masked auditory threshold and has presented data to indicate that in the actual situation multiple observers do not quite come up to predicted improvement. There are a number of reasons for this, the chief one being (as Schafer pointed out) that observers are rarely independent in their judgments, even if they are physically separated from one another. For one thing, if there is a noise background, they listen to the same noise. For another, an experimental test usually involves a long series of signals which are the same for all observers. A number of investigations have shown that human subjects have predictable biases in generating a series of responses. One study (7), in particular, has shown that a medium-strength signal is less likely to be reported if it follows a strong signal than if it follows a weak signal. This tendency also produces correlation among the observers (in an experimental test, at least). Also group morale, fatigue, etc. play a role.

It is obvious, therefore, that a theory of the multiple observer must consider the consequences of correlation among observers. It is easy to see that if the correlation is perfect, and all observers make exactly the same report at the same time, any one of them will be exactly as good as the group. The maximum gain comes when each observer's reports are independent of the reports of all the others (for a constant-strength signal, if negative correlations are not admitted). The theory must also examine the effects of individual differences among the observers with respect to their threshold and its standard deviation. Again it is obvious that if one observer is much more sensitive than the rest, he will do most of the reporting for the group, and the rest of the group could be discarded.

Another aspect of the problem—one that has not previously been considered—is of considerable importance. Human observers are more sensitive detectors if they are allowed to make false re-

[1] Research conducted at Lincoln Laboratory, Massachusetts Institute of Technology. The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

† Many people have contributed to the conducting of this experiment and to the development of the theory. There are a few we would like particularly to mention.

Dr. Bert F. Green has given us several very helpful suggestions, and we are particularly indebted for his technique of integrating the multivariate normal correlation surface.

Dr. J. C. R. Licklider supervised the calibration of the equipment.

The following people contributed to the design or construction of the equipment: Roy Sallen, Rudy Schreitmueller, Bob Silva, Josiah Macy, and Jack Flannery. We are very grateful for their help.

ports. That is, their threshold goes down, even when it is corrected for the amount of guessing implied by the false report rate. This gain is of little practical significance, however, unless some means can be found for bringing the false report rate back to an allowable level. One possible way of doing this is to allow each of a group of observers to make false reports, but to demand a high degree of agreement among members of the group before taking action.

The theory, therefore, should take account of (*a*) the variation between observers (as well as the variability of each of them around his own threshold), (*b*) the correlation between observers, and (*c*) the implications and consequences of false reports. The case in which there are no false reports by individual observers is simpler, and it will be examined first.

A. Independent Observers—No False Reports

1. *The Case of Two Equally Sensitive, Equally Variable, Uncorrelated Observers*

The usual model of a threshold assumes that some characteristic of the observer varies in time, rapidly enough that its positions on two successive trials are independent. This sensitivity is assumed to vary normally about a constant mean, usually with respect to the logarithm of signal strength. The psychophysical method of constant stimuli is assumed to display the integral of this distribution.

This model of the threshold works only when there are no false reports. A signal of zero strength is infinitely far down on a logarithmic scale, and if such a signal yields a finite probability of report, the area under the threshold distribution cannot be finite. However, if observers try very carefully to avoid false reports, experimental data gathered under many different conditions approximate this model fairly well, and we will use it for an initial examination of the multiple-observer situation.

For the assumptions made here the thresholds of the two observers are represented by two identical normal distributions. Their joint distribution can then be assumed to be represented by a normal correlation surface, with $r = 0$. To find the probability that one or the other of the observers will report a signal of strength S, we want the probability that one or both of the threshold values will be less than S, and this is one minus the double integral of the joint distribution from S to ∞ on each variable. This integral is a function only of its lower limit, hence as S varies from weak to strong, $f_c(S)$ plots the cumulative distribution. Since the means and σ's of the two distributions are assumed to be the same, S may be expressed as a deviation from the mean, divided by σ, and the cumulative distribution function of S is:

$$(1)\quad f_c(S) = 1 - \int_S^\infty \int_S^\infty \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dxdy = 1 - \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1 - \left[\int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx\right]^2 .$$

This last expression is the square of the integral of the univariate normal distribution function, and can be obtained easily from tables.

The probability density function is given by the derivative of $f_c(S)$ with respect to S. It is:

$$(2)\quad f(S) = 2 \frac{1}{\sqrt{2\pi}} e^{-\frac{S^2}{2}} \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx .$$

It should be noted that this distribution is the same as the distribution of the lower of two values chosen independently from the same normal distribution. The distribution is approximately symmetrical with a mean of $-.564$ and σ of .826.

$$f_e(S) = 1 - \int_S^\infty \int_S^\infty \cdots \int_S^\infty \left(\frac{1}{2\pi}\right)^{N/2} e^{-\frac{1}{2}\left[\left(\frac{X_1 - M_1}{\sigma_1}\right)^2 + \cdots + \left(\frac{X_n - M_n}{\sigma_n}\right)^2\right]} dx_1 \cdots dx_n \tag{3}$$

$$= 1 - \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1 - M_1}{\sigma_1}\right)^2} dx_1 \cdots \int_S^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n - M_n}{\sigma_n}\right)^2} dx_n \,.$$

As the number of independent, equal observers increases, the distribution of the lowest of the set of observers shifts down farther from the parent population. The cumulative distribution function will be the same form, but the exponent on the brackets will be equal to the number of observers. The means and σ's of these distributions have been tabulated by Hastings, Mosteller, Tukey, and Winsor (1) not only for the lowest, but for all orders in groups up to size 10. The mean for the lowest of three observers is $-.846$; for five, -1.163; and for 10, -1.539. The σ's for the distributions of the lowest decrease to .587 for $N = 10$.

2. *The Case of* N *Independent Observers: No Restrictions on Individual Means and Standard Deviations*

The formulas for the cumulative distribution of the best of N observers follow readily from the argument given above. The only difference is that we can no longer make the same transformation of S for all observers, since each is allowed to have a different mean and σ. If $M_1, \ldots, M_n$ represent the means of the observers and $\sigma_1, \ldots, \sigma_n$ the corresponding σ's, then the probability that at least one of N observers will respond to a signal of strength S is given by:

If we perform on each variable X_i, the transformation $z_i = (X_i - M_i)/\sigma_i$ we have:

$$f_c(S) = 1 - \int_{\frac{S - M_1}{\sigma_1}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}} dz_1 \cdots \int_{\frac{S - M_n}{\sigma_n}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z_n^2}{2}} dz_n \,. \tag{4}$$

This distribution can be plotted, with a little effort, from tables of the integral of the normal curve, for given values of the means and σ's.

In order to discuss the question of the set of means and σ's that maximizes the difference between the sensitivity of the group and the sensitivity of the best observer in the group, it is necessary to consider the concept of "best observer." The assumptions we have made so far allow two observers to have the same mean but different σ's, while both are normally distributed. If this condition were realizable, we would have the situation illustrated in Fig. 1. Observer A would be more sensitive to signals above the common mean, but observer B would be more sensitive to weak signals, and the combination would be very effective.

However, common sense indicates that this situation should be very difficult to obtain and experimental evidence confirms the opinion. In realizable situations, the σ's of individual subjects are not very different, and any considerable

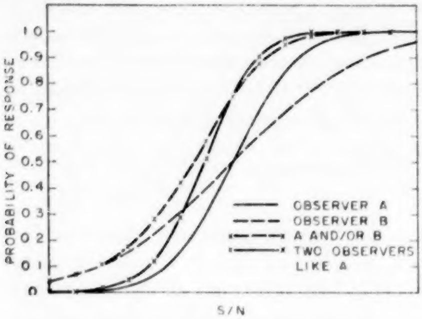

FIG. 1. Report probabilities for two observers with the same mean but different σ's. Curves marked with crosses show report probabilities for teams of two observers.

increase in σ is usually accompanied by a considerable departure from normality of distribution. As a consequence, the assumption of unrestricted variation of σ's is unreasonable, and can be abandoned without much loss of generality of application.

If we restrict our assumptions so that only the means are variable, and if we fix the lowest mean, then it can be seen from formula 4 that the probability of a report by at least one observer will be maximized if all have this mean. This may be demonstrated as follows. Suppose M_1 is the lowest mean. Then the difference between observer one (the best observer) and the group, for any value of S, is given by:

$$(5)\quad \int_{\frac{S-M_1}{\sigma_1}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}} dz \left[1 - \int_{\frac{S-M_2}{\sigma_2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_2^2}{2}} dz \cdots \int_{\frac{S-M_n}{\sigma_n}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_n^2}{2}} dz\right].$$

Now $M_2, \ldots, M_n$ are chosen to maximize 5, with the restrictions that $M_1 \leq M_2, \ldots, M_n$ (and all σ's equal). Since each of the integrals must have a value less than one, regardless of the value of S or of the M's, the expression in the brackets will always be positive. Consequently, the product of the integrals in the brackets must be minimized. Whatever the value of S, as M_i decreases, $(S - M_i)$ increases, and the value of the integral decreases. But since $M_2, \ldots, M_n$ cannot be smaller than M_1, they must all equal M_1 to maximize 5.

B. Correlated Observers—No False Reports

In the introductory paragraphs it was indicated that the major sources of correlation among observers are in the situation (the external noise, etc.). However, if we retain the assumption of a threshold that varies up and down in a normally distributed fashion, we must treat the situation as if the thresholds of the subjects actually covaried. We will later substitute what we believe to be a more adequate set of assumptions, but the assumption of covariation of thresholds is satisfactory for an introduction to the role of correlation. If observers make no false reports the picture is adequate. If variation in noise is the major source of correlation, then it makes no difference whether we consider the signal-to-noise ratio to be varying uniformly for all observers, or consider the individual thresholds to have a tendency to vary together.

The general method of deriving the cumulative distribution of the group of observers is the same as before. The probability that at least one observer will report a given signal is given by one minus the probability that none will report, or one minus the multiple integral of the normal correlational surface. However-

ever, when the correlations between variables are not zero, the expression for the normal correlation surface is formidable.

For economy of expression, let:

$\sigma_{ij} = \sigma_i^2$ when $i = j$ (this is the variance of the i^{th} observer)

$\sigma_{ij} = r_{ij}\sigma_i\sigma_j$ when $i \neq j$ (r_{ij} is the correlation between the i^{th} and j^{th} observer).

Letting i and j take all possible values from 1 to N, we have the variance-covariance matrix (a_{ij}). The inverse of this matrix is denoted by (a^{ij}) and the determinant of the inverse by $|a^{ij}|$. The multivariate normal correlation surface is:

$$f(x_1, \cdots, x_n) = \left(\frac{1}{2\pi}\right)^{N/2} \sqrt{|a^{ij}|}\, e^{-\frac{1}{2}\sum_1^N \sum_1^N a^{ij}x_ix_j}.$$

The a^{ij} in the exponent denotes the elements of the inverse of the variance-covariance matrix. The x's are expressed as deviations from their own means. A more complete treatment is given in Mood (2).

This expression is cumbersome to handle unless some simplifying assumptions are made. Without loss of generality the means and σ's of all observers may be assumed to be equal, since this represents a linear transformation of each variable, and can be compensated for by a change in the limits of integration. It does constitute a loss of generality to assume all intercorrelations of observers to be the same, but if all observers are assumed to be equal in sensitivity and in variability, the additional assumption of equal r's is not difficult to make.

We are willing to make the assumption of equality of intercorrelations, but even so, the expression is a very difficult one to handle. The usual methods of expanding in a series fail, since the series becomes completely unmanageable after a few terms. For $N = 2$ the problem is not too difficult, and values of the integral have been tabled by Pearson (5). Pearson (4) has also published a very ingenious expansion in powers of r for larger values of N, but the series blows up before it begins to converge. However, a method proposed to us by Dr. Bert F. Green allows a feasible, though tedious, solution. It will work in the general case, but is relatively simple only when we integrate up to the same value on each variable, which in this case means that we assume all observers to have the same mean and standard deviation.

The method consists of converting a correlational surface with n variables into a zero-correlation surface of $n + 1$ variables. For five variables, equally intercorrelated, and integrated up to the same z score on each, the development is as follows:

Take five normally distributed variables, $X_1 \ldots X_5$, each with mean zero and standard deviation of σ_x and all intercorrelations equal r.

Now take six normally distributed, independent variables $Y_1 \ldots Y_6$, all with mean equal to zero. Let the standard deviations of $Y_1 \ldots Y_5 = 1$ and of $Y_6 = \sigma_6$. Let

$$X_1 = Y_1 + Y_6$$
$$X_2 = Y_2 + Y_6$$
$$X_3 = Y_3 + Y_6$$
$$X_4 = Y_4 + Y_6$$
$$X_5 = Y_5 + Y_6$$

The correlation between X_i and X_j is the percentage of common variance, or

$$r = \frac{\sigma_c^2}{\sigma_x^2} = \frac{\sigma_6^2}{1+\sigma_6^2}\ ; \text{ and } \sigma_6^2 = \frac{r}{1-r}.$$

Now the integral of the normal correlational surface from $-\infty$ to $x_i = a$ for all i's is the probability that the following conditions will be met simultaneously.

$$\begin{array}{lcl} Y_1+Y_6<a & & a-Y_1>Y_6 \\ Y_2+Y_6<a & & a-Y_2>Y_6 \\ Y_3+Y_6<a & \text{or:} & a-Y_3>Y_6 \\ Y_4+Y_6<a & & a-Y_4>Y_6 \\ Y_5+Y_6<a & & a-Y_5>Y_6 \end{array}$$

Here we have six normally distributed, independent variables, and ask the probability that five of them are greater than a specified sixth. ($u_1 = a - Y_1$, $u_2 = a - Y_2, \ldots, u_5 = a - Y_5$) have a mean of a and σ of 1, but Y_6 has mean equal to zero and $\sigma = \sigma_6$. The desired probability is the probability density of Y_6 taking some value t, and all the others ($u_1, \ldots, u_5$) exceeding t, integrated over all values of t, or:

conscious sensation of tone, and that he reports only when he has it. His sensitivity is assumed to vary in time, so that the objective signal strength that leads to the experience is not always the same, but his probability of hearing extremely weak signals is infinitesimally small.

Within this framework it is difficult to account for the instances in which the observer reports that he hears a signal that was not physically present. It could be assumed that forcing a more liberal reporting attitude on the subject merely causes him to report positively in some fixed proportion of the trials on which he hears no tone. Consequently, a correction for guessing, based on the percentage of false reports, should yield the same threshold distribution that is given when there are no false reports. Preliminary results, however, showed that this was not the case.

$$P = \int_{t=-\infty}^{\infty} \left[\frac{1}{\sigma_6\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t}{\sigma_6}\right)^2} \left\{ \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u-a)^2} du \right\}^5 \right] dt$$

$$= \int_{-\infty}^{\infty} \left[\frac{1}{\sigma_6\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t}{\sigma_6}\right)^2} \left\{ \int_{t-a}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \right\}^5 \right] dt \cdot$$

This expression can be integrated numerically to give the probability that all the X's (correlated values) will fall below some specified value a. The process is tedious, but taking t in .1 steps and summing products seems to provide fair accuracy to the third decimal place.

C. The Interpretation of False Reports

All of the foregoing discussion is based on the implicit assumption that the observer either does or does not have a

It seems desirable to find a new framework in terms of which the threshold can be discussed. Referring to the conscious experience of tone adds little to the analysis of the threshold, and it seems better to view the report of the observer simply as a bit of behavior, determined jointly by the instructions, the signal, and his own internal variability. The details of this analysis are so closely linked to the data to be presented that further discussion at this point is unprofitable.

II. THE EXPERIMENTS

The plotting of the relationship between signal strength and relative frequency of report of the signal requires great quantities of data, gathered by the method of constant stimuli. Moreover, to get even a relatively stable result from instructions, particularly with naive observers, it is necessary to keep the observer informed of his results. Both of these requirements are difficult to meet without the aid of automatic machinery.

The main experiment was designed to permit five physically separated observers to listen simultaneously to the same 800-cycle tone (masked by a broad-band noise) and to report independently whether or not they heard it. Twenty-three groups of five men each participated in the experiment. Twelve of these groups were instructed to indicate that they heard the tone only when they were quite certain of it. The other eleven groups were instructed to report whenever they *thought* they heard the tone. Thus, data were gathered for units of five observers under two different attitudes toward responding.

A. Equipment

1. *General*

Figure 2 shows a block diagram of the apparatus for presenting signals and recording responses. The unique feature of this apparatus is the automatic signal selection. The multiple-tap attenuator had 16 output leads, each carrying the signal, but with a different degree of attenuation. These ranged from zero attenuation through 14-db attenuation in one-db steps ("db," of course, stands for *decibel*). The sixteenth output was tied directly to ground and carried no signal.

The 16 signal strengths were fed into a binary fan (see Fig. 2A) which selected one of the 16. The selection was controlled by the pattern of energization of the four relay coils. This pattern was determined by the pattern of holes punched in a teletype tape. The 32 symbols transmitted by teletype may be coded into a pattern of five holes punched (or not punched) in a line across the tape. Since only 16 signal strengths were used, only four of the five rows of holes were necessary. As the tape advanced through the teletype reader, the reader transmitted the pattern of holes to the four relays. For example, if there were holes in the tape in positions 1, 3, and 4, the corresponding relays were closed, and the signal at 13-db attenuation was selected. A sequence of signal strengths was punched into the tape, and the tape was stepped through the reader by an impulse coordinated with the signal timer.

The responses of the observers were also punched into a teletype tape. A teletype perforator was modified by mounting solenoids under the five keys that corresponded to a single punch in one of the five rows. The observer was given a response box with a spring-return toggle switch, and told to push the switch on the trials on which he heard the tone. Each observer controlled one of the five solenoids, so that a performance could be scored by counting the holes in one row of the response tape. In normal operation the teletype perforator advances the tape as soon as any key is depressed, punching out the code corresponding to that key as it does so. To keep the responses to a given signal lined up, the normal perforator circuit was disconnected and a

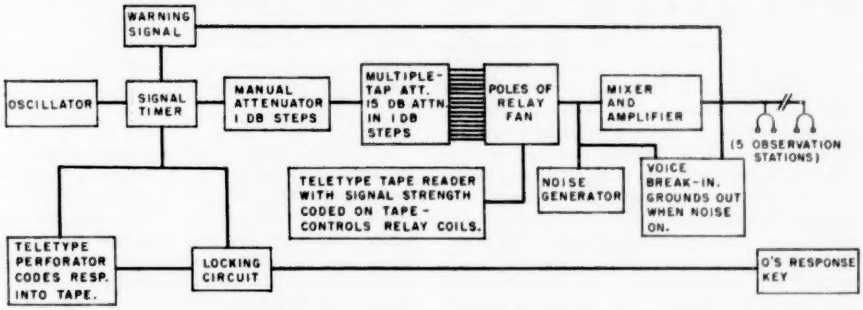

FIG. 2. Block diagram of apparatus for producing signals and recording responses.

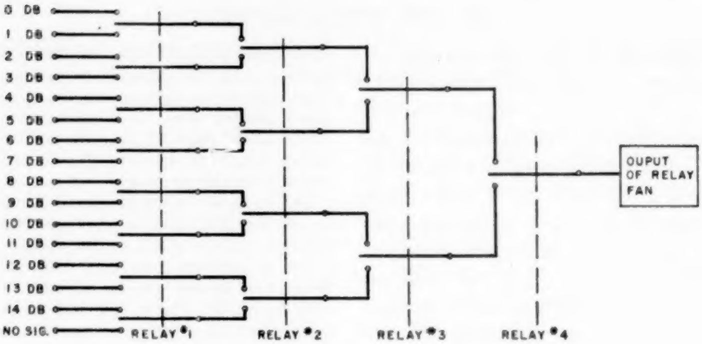

FIG. 2A. Schematic diagram of binary relay fan.

pulse from the signal timer advanced and punched the tape. The observer's response key actually controlled a locking relay, which, in turn, kept the appropriate solenoid energized until the tape was punched. The same impulse that actuated the perforator cleared the locking relays.

The observer's response key was a double-throw switch. For the main experiment, only a yes-no response was required and the two poles were connected together so that the switch could be operated in either direction with the same result. However, in some experiments, four categories of response were required (corresponding to different degrees of certainty that a tone had been presented). For this condition the response-box wiring was changed so that one box controlled two rows of holes on the response tape. The four categories of response were indicated by the observer by throwing the switch both ways, to the right, to the left, or neither. Only two observers could be run simultaneously under these conditions.

Reading the response tapes took additional apparatus. Counting the number of holes in one row of the tape gave the total number of responses, but this total had to be broken down by signal strength to be useful. The information on each response tape had to be broken down into 80 separate counts—the number of times each of five observers responded to each of 16 signal strengths. To do this took a panel of 80 electromagnetic counters (5 columns and 16 rows). In order to record the counts by signal strength, two teletype readers were run synchronously, one reading the signal tape and the other the response tape. The signal-tape reader selected the row of counters to be activated, and a hole in the response tape was recorded on the counter in the corresponding column. The readers were advanced every second, so that a tape containing responses to 180 signals could be scored in three minutes. This device made it feasible to score the performance after each trial and inform each observer of his results before the next trial.

Two other scores were needed. For the evaluation of the group as a joint detection unit, it was necessary to count, by signal strength, the number of times 0, 1, 2, 3, 4, or all 5 observers responded to a given signal. From these counts curves could be plotted showing, for each signal strength, the percentage of times at least one, at least two, etc. observers responded. A special relay circuit, fed by the response tape, made this count possible.

The other count was for the intercorrelation of observers. This, again, had to be made by signal strength, since the correlation is meaningful only if signal strength is partialed out. The count was also made separately for each pair of subjects. Ten different pairs can be drawn from five observers, so each response tape contributed data toward 160 different correlations. The counts that had already been made provided the marginal totals of a two-by-two correlation table for each pair of observers. The only additional information needed was one of the cell frequencies. This was obtained with the help of another relay circuit, which responded to coincidence of punches in a specified pair of rows on the response tape.

2. *Stimulus Data*

a. Timing of signal. For all experiments the signals were separated by a uniform period of five seconds. The duration of the signal was .41 second. The noise was on continuously for the whole period of the test, and each signal (or blank) was preceded by about one second by a warning signal. The warning signal was a sharply damped oscillation that was clearly audible above the noise.

b. The signal. The signal was an 800-cycle pure tone, generated by a General Radio oscillator. Its intensity, for the loudest signal used, was —37.3 db re 1.0 volt. This measurement was made at the earphones, with a Ballantine Voltmeter.

c. The noise. The noise was generated by a 6D4 gas tube, amplified. The spectrum of the noise generator, as measured at the earphones by a Hewlett Packard wave analyzer set for 135 cps half bandwidth, was flat to within one db from 400 to 2500 cps, with a drop of 10 db from the peak down to 100 cps, and a drop of about 12.5 db from the peak up to 10,000 cps.

The noise intensity was measured for a band from 500 to 2400 cps. An amplifier with a high input impedance and a band-pass filter were connected in parallel with the earphones, and the output of the filter was measured with a Ballantine Voltmeter. The rms voltage at the output of the filter was —28.0 db re 1.0 volt. This value represents a correction of the actual Ballantine reading for the fact that the noise peaks were clipped, and a correction for the fact that noise, rather than a sine wave, was being measured. The noise level in a one cps band was, therefore, —60.8 re 1.0 volt, for the 1900 cps band from 500 to 2400 cps.

The signal-to-noise level of the loudest signal was —37.3 — (—60.8) =23.5 db.

d. The earphones. The earphones used were PDR-8's. Their frequency response curves were obtained for a 1 volt constant input with the earphone feeding into a 6-CC coupler. The resulting sound pressure readings were corrected (3) for the earphone cushions that the observers wore.

The sound pressure level was fairly uniform over the range from 100 to 7-8000 cps. Of the ten individual phones used, nine varied less than 4 db in the range from 500 to 2500 cps. The one exception had a sharp peak at 2000 cps, followed by a dip at 2500 cps. This range was 15.5 db, but because of the frequencies at which the deviations occurred, the noise level as calculated above should apply even in this case. None of the phones showed any peculiarities in the immediate region of 800 cps.

At 800 cps the sound pressure levels varied from 100.0 to 107.0 db re one microbar for a 1 volt input. This maximum deviation was between the right and left phones used by observer No. 2. The other pairs were better matched, differing by 1.5, 1.0, 4.0, and 0.5 db.

The loudest signal used was —37.3 db re 1 volt. Hence the sound pressure level for this signal was between 62.7 and 69.7 db re one microbar, depending on the phone used. The noise, in terms of energy per cycle, was 23.5 db below this.

e. The range of signals. The automatic attenuation was supposed to cut the signal intensity by 1-db steps, over a range of 15 db. It failed slightly in doing this. No single step was off by more than 0.1 db, and the maximum cumulative error was 0.3 db. The actual intensities, expressed as decibels attenuation from the loudest signal, were 0, 1, 2, 3, 4, 5, 6.1, 7.1, 8.2, 9.3, 10.2, 11.2, 12.2, 13.1, and 14.2. All graphs expressing results are corrected for these errors.

B. Observers

The observers were all enlisted military personnel from Fort Devens, Massachusetts. They were brought in each morning and returned in the evening, and were available for testing for five or six hours.

Twenty-three groups of five men each participated in the main experiment. Of these 115 observers, 103 were different individuals. In seven instances, men were erroneously returned for a second test period. In addition, one group of five men was intentionally returned for a number of days for another purpose. On their first day they ran under the standard conservative instructions. On the second day they ran under standard liberal instructions. In no case was there evidence of a practice effect from the first day's testing, so these 12 repeat observers were treated in the same way as the rest of the men.

On the whole, the men were remarkably cooperative. Conditions in the listening rooms were far from ideal, particularly for the earlier groups. Some of these were run in unventilated cubicles on a hot summer day. Often the men reported that they were short on sleep, but there were only about a half dozen known instances of observers falling asleep in the listening room. The experimenter made a practice of checking the response tape to correct this situation if it arose. Data from a 24th group were discarded because the members could not be kept awake, but in the other cases, no correction of the results was possible. This factor thus represents a slight source of error.

The Army General Classification Test scores of the observers were obtained from Fort Devens when possible. The average of the known scores was 109—slightly above the mean of the military population.

C. Procedure

When the group of men arrived in the morning, they were given a brief explanation of the purpose of the experiment, followed by detailed instructions on their task. They were then given a demonstration of the sound of the noise and the signal in the noise. At the end of the demonstration all men were questioned to insure that they had recognized the signal, and knew what

to listen for. They were then given a practice series, made up in the same manner as the test series. The practice series was scored and the results explained to the men before the experiment proper began. There is evidence that two men still failed to understand (in the early part of the experiment at least) what was expected of them. Since the interest was in the performance of the whole group, the data from these men are included in the totals. Their performance curves for the whole day were not very atypical.

The experiment itself was made up of listening periods of about 15 minutes each. Each listening period was made up of 180 signals (10 of each of 15 different strength signals, and 30 blanks). It was possible to get in 6 to 9 listening periods in the course of one day. One group had 6 periods, 2 had 7, 15 had 8, and 5 groups had 9 periods. Thus the typical observer responded to 1,440 signals, 80 at each of the signal strengths and 240 blanks.

The stimulus tapes were made up in a random sequence, subject only to the restriction that 10 of each signal and 30 blanks occurred on each tape. Every tape the observers heard had a different sequence, but the same set of tapes was used for all groups. Preliminary experience showed that the first few responses on each tape were apt to be unreliable, so each tape began with four extra signals, including one loud one. Observers responded to these four, but the responses were not scored.

Every response tape was scored, and a general picture of the results was given to the observers before the next tape was run. It is believed that this procedure was, in large part, responsible for the high degree of cooperation we obtained from the observers. Typically, they watched the scoring and it was not unusual for strong competition to develop in the group. The immediate knowledge of results also helped considerably in stabilizing the attitude of the observers toward reporting.

The portion of the instructions designed to induce a conservative or liberal attitude toward the reporting of signals is given below. Additional caution or encouragement was given during the course of the day, when it was indicated by the results.

Conservative groups:

Keep in mind that it is important for you to be sure you hear the tone. In many cases there will be no tone, and you will be making a mistake if you indicate that you hear one. None of the tones is really easy to hear, but don't push the switch unless you are sure that you heard the tone. If you are in real doubt as to whether or not you heard a tone, assume that there was none.

Liberal groups:

Keep in mind that all of these tones are hard to hear, and that you will rarely be absolutely sure that you heard something. But if you think you heard something, probably you did—report it. If you are very sure that you didn't hear anything—then don't touch the switch.

In addition to the main experiment, two groups of five men each were tested under special conditions. These men were asked to respond with one of four categories of response, instead of the two categories (yes or no) required in the main experiment. These four categories were verbalized as follows:

1. Certain that there was a tone. This corresponds to the "yes" category used by the conservative groups.
2. Think there was a tone. If a response fell in either category 1 or 2, it was assumed to be comparable to the "yes" category of the liberal groups.
3. Didn't hear it, but guess it was there.
4. No tone.

Categories 3 and 4 were intended to break down the "no" category of the liberal groups.

III. RESULTS

A. Cumulative Response Curves

1. *Curves for All Observers*

The quantity of data gathered makes it possible to get a good picture of the report rate as a function of signal-to-noise ratio. Figure 3 gives this plot for all observers. The conservative curve is based on the performance of 60 observers, with a total of 4,900 judgments on each signal strength, and 14,700 judgments on blanks. The liberal curve is based on 55 observers, with 4,350 judgments for signals and 13,050 for blanks. Each judgment was given equal weight, although the number of listening periods for an individual varied from six to nine.

The curves approximate a normal ogive, but the approximation is not very close, even for the conservative curve. The major deviation from the ogive lies in the fact that the false report rate is not

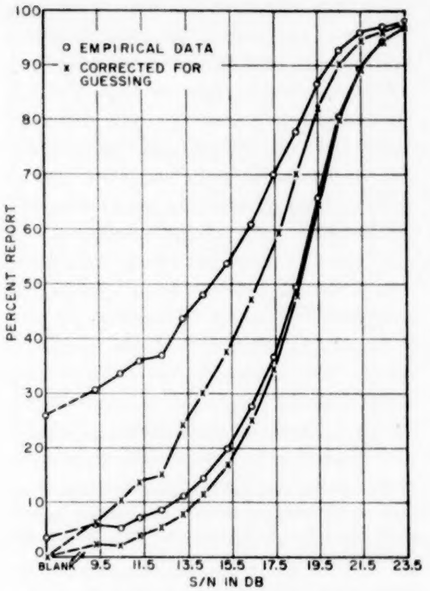

FIG. 3. Averages for all observers in main experiment. Lines marked with crosses are corrected for guessing. The top two lines are the raw and corrected averages for observers given liberal instructions; bottom two lines are for conservative observers.

zero. However, this deviation extends up past the middle range of signals. When the two curves are plotted on normal probability coordinates they still tend toward the shape of an ogive. The upper branch of the curve (above 18.5 db) is an approximation to the normal distribution in the case of the conservative groups, although the two loudest signals were reported somewhat too infrequently for the distribution to resemble the normal very closely. The same is true of the liberal observers, to a greater extent. These curves are known to include some sources of error—partial equipment failures, drowsy or uncooperative observers, etc. However, it was impossible to discard data for individuals in considering the effectiveness of the group, so these curves are presented for comparison with the curves for the group considered as a multiple detection unit.

Each of these curves shows one marked deviation from a smooth line. The liberal curve has a dip at 12.5 db and the conservative curve at 10.5 db. It is difficult to give a meaningful statement of the significance of these deviations, but at least one of them appears to be within the limits of sampling variability. To test this we took individual percentage reports for the questionable point and for the point on either side. Each percentage was subjected to the arc sine transformation, and the variance of individuals was computed for each point. The variances for the three liberal points were tested for homogeneity by Bartlett's test, and so were the three conservative variances. Both sets were found to be homogeneous, so the estimate of the variance of an individual point was based on the variance of the set. The expected percentage in each case was estimated graphically, and converted to an arc sine. The difference between the estimated and the observed values was divided by the appropriate standard error of the mean to yield a t value. For the dip in the conservative curve, this t was .96 and for the liberal curve it was 2.00. For 54 df the 5 per cent level of t is 2.01, so that even the dip in the liberal curve is not quite up to the conventional level of significance. However, these are conservative estimates of significance, since they ignore the correlation between observed and estimated percentages.

There is no reason to suspect the apparatus in the case of either of these dips. The groups were run on alternate days, so any apparatus failure too intermittent to be detected by other means should show up in both groups.

The correction for guessing. If the

liberal observers attain their higher report rates by guessing, it should be possible to detect it from the data. We assume that an observer either hears the signal or he doesn't, but that if he doesn't hear it, he has a certain probability of reporting anyway. If this is the case, by instructing the observers to be liberal in their reporting, we have merely increased the probability of reporting when nothing is heard.

Under these assumptions, when no signal is presented, none can be heard, and the number of false responses gives us the probability of a guess. Let P_t be the true proportion of signals heard, and P_g be the proportion of guesses. Then $(1-P_t)P_g$ will be the proportion of reports attributable to guesses. If P_o is the observed proportion of responses, then $P_t+(1-P_t)P_g=P_o$; and $P_t=(P_o-P_g)/(1-P_g)$. P_t is the corrected proportion of reports.

These corrected percentages are shown in Fig. 3. If the assumptions we made above were true, these corrected curves under liberal and conservative attitudes should be the same. Since they are not the same, the additional reports made under the liberal attitude represent something more than a guess.

2. *Curves For Selected Observers*

Figure 4 gives a better picture of what the cumulative response distributions should be. It is plotted on normal probability coordinates so that deviations from normality may be seen easily. It was pointed out in the previous section that the average curves for all observers were known to contain some sources of error. An attempt was made to eliminate the more gross sources of error. It was felt that any observer should have been able to report 95 per cent or more of one of the louder signals if he had been awake

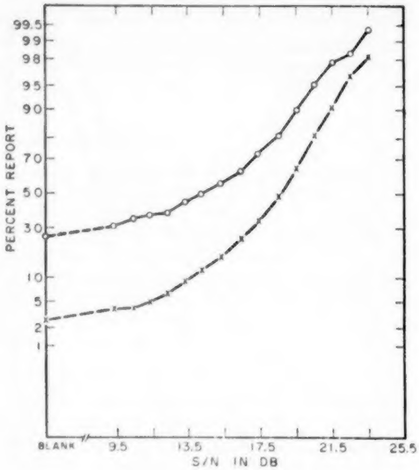

FIG. 4. Average performance of *selected* observers, plotted on normal probability coordinates. Top line is liberal observers, bottom line conservative.

and alert throughout the experiment. Therefore, we discarded the data of any observer in either group who failed to reach the 95 per cent report level for at least one signal strength. This criterion eliminated 6 from the conservative group and 7 from the liberal group. This criterion may seem too stringent for the conservative group, and the question might be raised as to whether we are discarding observers who lapsed into another state (and hence do not represent the same population as the rest) or whether we are capitalizing on error and discarding the extremes of a continuous distribution. Actually there was a strong tendency in both groups for those who fell below 95 per cent to fall well below it. A few of the observers who were discarded probably represent extreme cases of insensitivity to the masked signals. However, we feel that inattentiveness or actual sleep was responsible for most of these exclusions.

In order to make the selected groups

more homogeneous, an additional cut-off was established for false report rate. Conservative observers with false report rates of 5 per cent or higher were discarded (an additional 13 observers). Liberal observers whose false report rates fell below 15 per cent (two cases) or above 35 per cent (two cases) were also discarded. This left 41 observers in the conservative group, and 44 in the liberal group. Unfortunately, these criteria do not eliminate all sources of error. For example, the dip in the liberal curve at 22.5 db is almost certainly due to occasional equipment failures at this signal strength. Most of this dip comes from one or two groups, and the records showed an extreme tendency for all five observers to miss the same signal. However, it was felt that any further selection of cases on the basis of the records themselves could lead to an unrealistic picture of the process under examination.

The resulting curves (Fig. 4) are not very different from the total averages. The conservative curve is sharpened, that is, it is lower for the weak signals and higher for the strong ones. This, of course, is to be expected from the nature of the selection process. The upper branch of the conservative curve for selected observers more nearly approximates the normal, but still shows a falling off for the strongest signal. The selected liberal observers have the same false report rate as the unselected group, but the rest of the curve is raised. This curve also shows a better approximation to normality in its upper half, but still falls off for the strongest signals.

B. The Performance of Multiple Observers

Cumulative response curves were computed for the group as a detection unit. These curves show, for each signal strength, the percentage of times that at least one observer responded to a particular signal; the percentage of times at least two responded; and so on up to the percentage of times all five of the observers responded. All the reports that go to make up the percentage report for the "two-or-more" curve also go into the curve for "one-or-more." Thus the five curves may have the same ordinate, but they cannot cross one another. These curves are plotted as parts of Figs. 16 and 17, Fig. 17 for the observers working under a liberal attitude toward reporting, and Fig. 16 for observers instructed to report only when they were certain they heard the signal (conservative attitude).

If the reports that go to make up these curves were used in a detection situation, it would be necessary that the false report

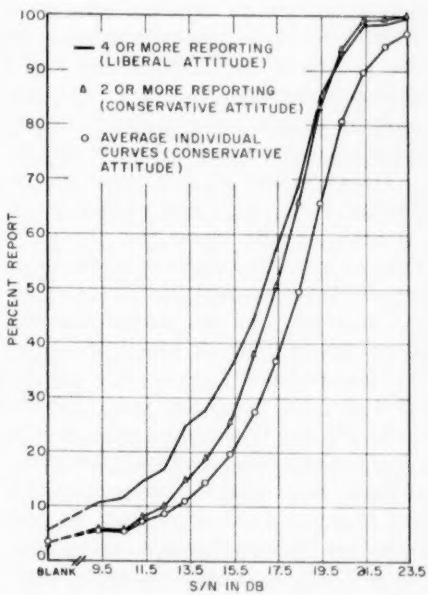

FIG. 5. Comparison of average individual performance with group performance under liberal and under conservative instructions.

rate be low. With observers working under a liberal attitude toward reporting, for example, it would probably be necessary to adopt a criterion of four or more observers reporting. Even this criterion gave a false report rate of 5.7 per cent. The requirement that all five observers report the signal brings the false report rate down almost to zero for the liberal groups, but also has a severe effect on the report of stronger signals.

When false report rates are held constant, there is little difference between the effectiveness of the groups working under the two different attitudes. Fig. 5 shows a comparison of the curves for "four-or-more" reporting under the liberal attitude, and "two-or-more" reporting under the conservative attitude. These curves have approximately the same false report rate, and the cumulative response curves are quite similar. The liberal groups were somewhat more sensitive to weak signals, but this may be due in part to the fact that they had a slightly higher false report rate. The conservative groups were actually slightly superior in detecting the stronger signals.

The curve for "all five" observers reporting under the liberal attitude has a .15 per cent false report rate. To get a lower rate for the conservative observers, we also have to adopt a criterion of all five reporting. In this comparison the liberal group is clearly better. However, the "four-or-more" (.17 per cent) and the "three-or-more" (.78 per cent) criteria yield very low false report rates for the conservative groups. If we compare either of these curves with the comparable liberal curve (all five reporting), we see again that the liberal groups are superior on weak signals, inferior on strong ones.

Figure 5 also compares the two group curves with the average curve for individual observers working under the conservative attitude. The false report rate for the individuals is almost identical to that of the conservative groups (two or more observers reporting) so that a direct comparison is possible. If the effectiveness of the group is measured in terms of the strength of signal which is reported 50 per cent of the time, then the group is not very superior. This gain is only slightly more than one db. However the curves show that this comparison in terms of the conventional threshold does not tell the whole story. The group makes a sharper distinction between signal and no signal than the individual does. If we compare the two on the strength of signal that will be reported 95 per cent of the time, the group is two db better.

C. Intercorrelations among Observers

In the main experiment, five observers listened to, and reported on the same signals. In the introduction to this report it was shown that the effectiveness of the group was dependent, in part, on the tendency for the observers to report, and to miss, the same signals. More than this, the correlations between observers yield an estimate of the proportion of the variance common to all observers. Since the observers were physically isolated from one another, there is no way that one of them could have gained knowledge of the response made by the others to a given signal. As a source of common variance, then, we are left with variation outside the observers —i.e., the masking noise. It is possible that other, and unintended, noises occasionally penetrated the sound-shielded listening booths, the earphone cushions, and the masking noise. These external noises may have made a small contribution to the common variance. Momentary variations in the masking noise,

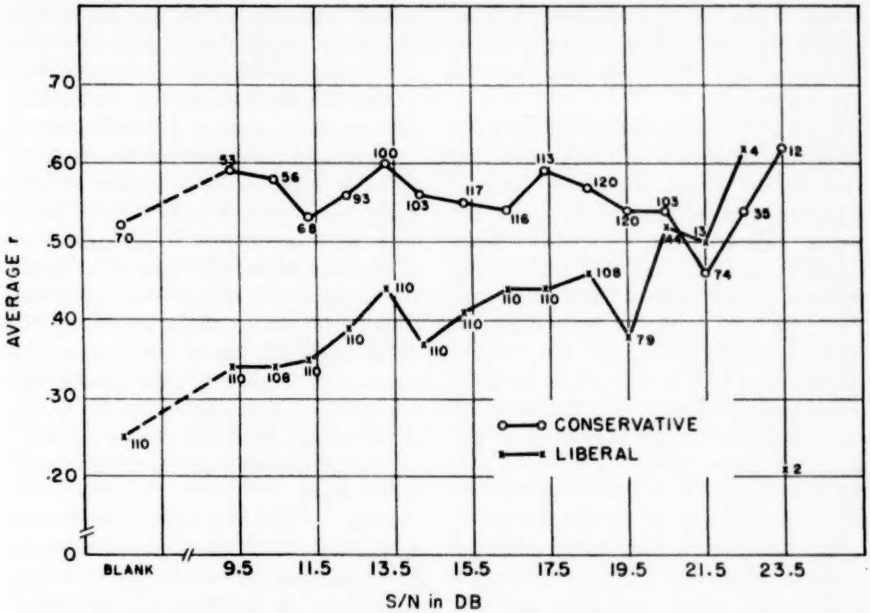

FIG. 6. Average intercorrelations of observers working under conservative instructions (top line) and under liberal instructions (bottom line). Alongside each point is the number of correlations making up the average (120 possible for conservative curve; 110 for liberal).

however, seem to be a more important source of common variance. Other secondary sources include those mentioned in the introduction, fatigue, sequence effects, etc.

The correlations between observers are meaningful only if the effect of signal strength is partialed out, since there is a strong tendency for all observers to report the strong signals and to fail to report the weak signals and the blanks. For this reason, correlations were computed by signal strength.

The data were given in the form of a dichotomy, report or no report, and the tetrachoric correlation coefficient was used. Separate correlation coefficients were computed for each of the ten pairs in each group of five observers, so that each group yielded ten correlations for each of 16 signal strengths. Each individual correlation was based on 60-90 observations for signals, and 180-270 for blanks. The correlations for a given signal strength were averaged by converting each r to Fisher's Z, averaging the Z's, and converting back to r. The resulting average correlations are plotted by signal strength in Fig. 6. Separate plots are given for liberal and conservative observers.

The nature of the data was such that correlations could not always be computed. When the signals were strong, individual subjects often reported every signal. In this case, the correlation is indeterminant. In other instances each of two observers missed one or two signals. If the missed signals were the same for both observers, the correlation was $+1$;

if not, it was -1. Perfect correlations obtained this way can also be considered indeterminant. They were not included in the average.

Thus the correlations for the strongest three or four signals in both groups, and also for the weaker signals in the conservative groups, tend to be quite unstable. Not only are there fewer intercorrelations to average, but each of the individual correlations tends to be unstable because of the extreme marginal split. However, feeling that an unstable estimate was better than none, we computed and averaged every determinable correlation, no matter how extreme the marginal split. The graphs in Fig. 6 show, alongside each point, the number of correlations making up the average.

The correlations between members of conservative groups fell between .5 and .6 for most signal strengths, and the question arises as to whether or not these correlations can be assumed to be drawn from the same population. An accurate answer to this question is extremely difficult to obtain, but we can make a crude estimate. The standard error of the tetrachoric correlation is given as:

$$\sigma_{r_t} = \frac{\sqrt{pqp'q'}}{z_x z_y \sqrt{N}} \sqrt{(1-r^2)\left[1-\left(\frac{\sin^{-1} r}{90^\circ}\right)^2\right]}$$

where p and q represent the marginal proportions of one variable with z_x as the ordinate of the normal curve corresponding to that proportion. The corresponding proportions and ordinate of the other variable are p' and q', and z_y.

The average r is approximately .55. We can make an approximate test of the hypothesis that the various average r's are drawn from a population with $r = .55$ by the following procedure. We assume that the two marginal distributions are the same, and take the data for these values from Fig. 3. We simplify further by taking a single value of N (the number of cases on which a single tetrachoric r is based). For convenience, we used 81. The standard error of a tetrachoric r of .55 was then computed for the marginal splits at each signal strength. To estimate the standard error of the mean correlation we divided by the square root of the number of correlations on which the average was based. The individual tetrachoric correlations are not normally distributed, but the distribution of the mean of 50 or more of them should tend toward normality, so that the deviations of the means from the over-all mean of .55 were evaluated in terms of the normal distribution.

Three of the 16 average correlations deviated beyond the 5 per cent confidence limits so established. The value of .60 at 13.5 db exceeded the confidence limit of .597. At 17.5 db, r was .59; the confidence limit is .576. At 21.5 db r was .46; the lower confidence limit is .494. If we had real confidence in these limits, we should have to reject the hypothesis that these correlations came from the same population, since, if the probability is .05 that any one will fall outside the limits, the probability (from the binomial expansion) is .047 that 3 or more out of 16 will exceed the limits.

Possible sources of error in this estimate are so great, however, that it can be taken as little more than an estimate of the order of magnitude of the error. The correlations are not all independent as assumed by the test. There is reason to believe that the groups varied among themselves more than might be expected on the basis of the same sort of calculation. If we look at the variability of groups around the average value of .55, the extreme deviations from .55 do not look very great. At 17.5 db, for example,

seven groups averaged above .55, one averaged .55, and four were below. At no signal strength were there more than nine of the twelve groups above or below the average r of .55.

It seems reasonable to conclude that we have no good reason for rejecting the hypothesis that the correlation between observers is independent of signal strength when observers adopt the conservative attitude. The conclusion can only be regarded as tentative, but at least there seems to be no evidence in these data for any trend.

Even the evidence for absence of a trend is doubtful. When we exclude from consideration those observers having a false report rate exceeding 5 per cent, and also those who fail to reach 95 per cent report on any of the strong signals, we find some indication of a trend in the figures. For this selected group of conservative observers, the correlations on the weaker signals were higher than those on the stronger signals. In making up this selected group of observers, 19 men were discarded—13 for a high false report rate, and 6 for failing to reach 95 per cent report on any signal. The discards were so arranged in the groups that there were 60 possible intercorrelations among the selected observers. The average intercorrelations, together with the number of individual correlations on which the averages were based, are as follows: .64(31); .72(24); .66(24); .64 (33); .59(46); .62(48); .58(50); .58(57); .56(58); .62(57); .58(60); .54(60); .56(57); .52(38); .60(14); .46(1). These correlations are listed in order of ascending signal strength, with the first one being the correlation when no signal was presented.

It may be seen from these figures that practically all of the correlations were increased with respect to those from the unselected groups. This was to be expected, on the grounds that the selection procedure eliminated those observers who were inattentive, or who failed to follow instructions (too liberal in reporting). The relatively greater increase on the weak signals was not expected, however. It might be argued that our selection on the basis of false report rate eliminated those whose error variance in reporting blanks happened to be high. This would explain the relatively large increase in the correlation for the blanks, but not for the weak signals. There was no explicit selection on the basis of response to these, so that our selection procedure should not have systematically increased these correlations.

For the liberal groups, Fig. 6 shows that the average intercorrelation was not independent of signal strength. There is a definite trend in the correlations, with those for blanks and weak signals being lower than those for the stronger signals. The data for selected liberal observers show the same thing. The selected observers average .07 higher, but this increase is uniform (within reasonable sampling limits) for all signal strengths.

D. Rate of Information

If we consider the individual observer as a device for decoding signals into reports on the presence or absence of a signal, we can arrive at another method of evaluating the effect of the attitude taken by the observer. If the observer adopts a liberal attitude toward reporting, this is analogous to turning up the gain on the last stage of an amplifier system. If the signals being amplified are weak, amplifier noise will be increased along with the signal, and the result may not be helpful. In the case of the observers, more signals are reported, but more blanks are reported as signals, too. The

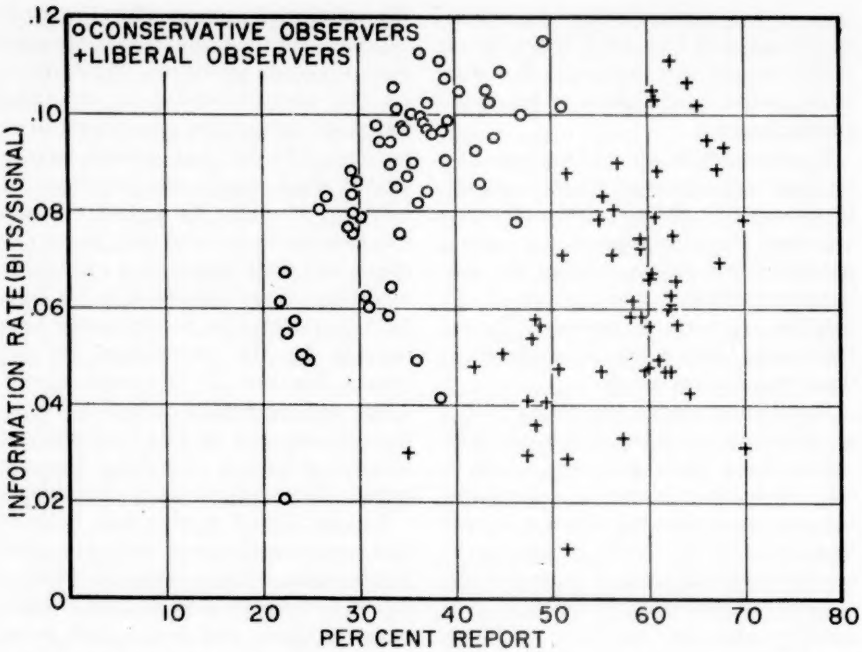

FIG. 7. Individual values of information transmitted per signal plotted against percentage report (signals and blanks).

question arises as to whether or not the increased report rate increases the amount of information being transmitted by way of the observer.

Shannon's (8) formula for rate of transmission in the presence of noise was applied to these data. [Rate = $H(x) + H(y) - H(x, y)$.] It should be noted that the absolute values of the calculated rates have very little meaning. These values depend on arbitrary factors, such as the percentage of signals (as opposed to blanks) used in the experiment, and the strength of signals used. The maximum rate of transmission under the conditions used (83.3 per cent signals and 16.7 per cent blanks) was .65 bits of information per symbol. The actual rates of transmission by the observers were far below this, since many of the signals were quite weak. However, the statistical characteristics and the strengths of the signals were the same under all experimental conditions, so that a comparison of conservative and liberal subjects seems justified.

For these calculations signals of all strengths were lumped together. A fourfold table was constructed for each observer, showing the percentage of times he reported a signal when there was and when there was not a signal actually given, and the corresponding percentages for his failures to report. The marginal totals on one side gave then the percentage of signals and blanks (the same for all observers), and on the other side the observer's percentage of reports. The formula given above was applied to these tables.

Figure 7 summarizes these calculations for the data from the main experiment. It shows rate of information in bits per signal plotted against the percentage of times the observer reported (to a signal or a blank). Each dot represents an observer working under the conservative attitude, and each cross a liberal observer.

Inspection of the plot shows that the liberal observers transmitted somewhat less information than did the conservative observers. The mean rates were .064 and .085 respectively.

On the other hand, it may be seen that within each group the information rate is positively correlated with the percentage of responses. Particularly in the case of the conservative observers, those with high report rates transmit more information. However, this correlation is due in part, if not in full, to differences in sensitivity of the observers. The men with high report rates are those who most effectively discriminated signals from noise. If it were possible to compare only observers equal in sensitivity, much of this correlation would probably disappear.

There is one fairly strong indication that this is the case. Although over-all report rate and false report rate are positively (but not linearly, see Fig. 8) correlated, there is no correlation between information and false report rate. Figure 9 shows this plot for liberal and conservative observers, and indicates that, within each group, the correlation is zero or slightly negative. This suggests that those observers in the conservative group who adopted a more liberal attitude (and hence increased both report rate and false report rate) did not tend to transmit much more information.

It may be that the observer who is too cautious in reporting signals trans-

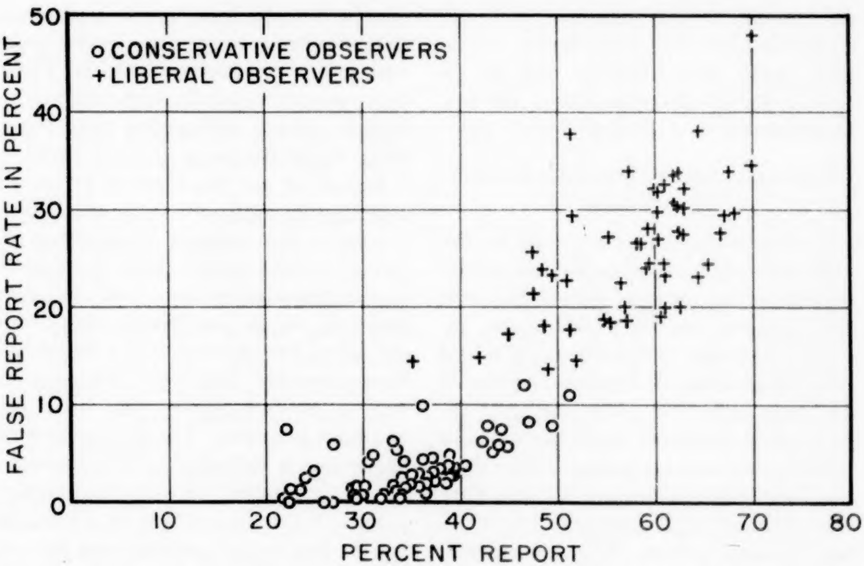

FIG. 8. Individual values of false report rate (percentage report of blanks) plotted against percentage report (signals and blanks).

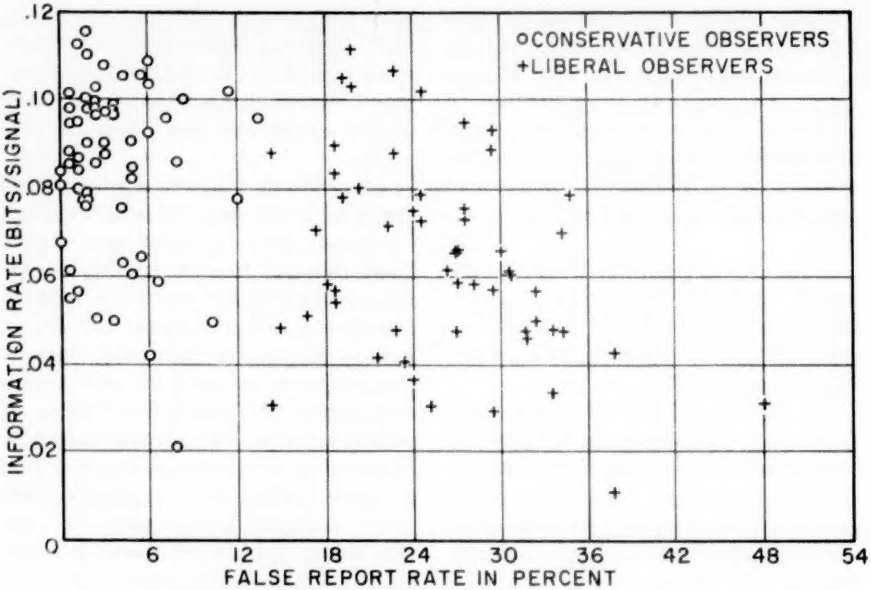

FIG. 9. Individual values of information transmitted per signal plotted against false report rate.

mits less than his maximum amount of information. These data give no direct indication of it. They do indicate, on the other hand, that when the attitude toward reporting becomes liberal, the rate of transmission of information drops.

E. Special Groups—Four Categories of Response

Two special groups were tested to provide more information on the interaction of cumulative response curves and attitude toward reporting. With the exception of minor differences in the initial training procedures, these two groups of five men each were given the same experimental treatment. Both groups were tested over a five-day period, rather than the usual one-day period, so that they were relatively well practiced. The members of both groups showed every evidence of interest in the experiment, so that their results are judged to be not only more stable than those of the other groups, but also more nearly representative of what a cooperative and competent group of observers can do. These men were unselected with respect to auditory acuity, and seemed to be representative of the larger sample tested.

Instead of the usual yes-no categories of response, these observers were asked to make a four-category judgment in response to each signal. They were asked to indicate whether they were sure they heard the signal, they thought they heard the signal, they guessed that a signal had been presented although they couldn't hear it, or they guessed that no signal had been presented. The first category of response was intended to be equivalent to the responses of the conservative groups. A response falling in either the first or the second category was equivalent to a response by one of the liberal groups. The sum of responses in cate-

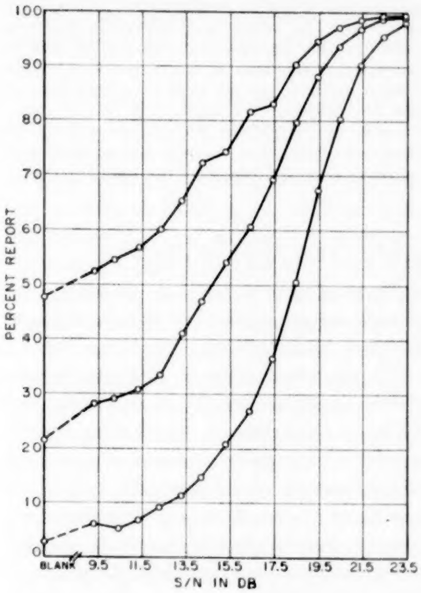

FIG. 10. Percentage report of observers responding in four categories. Bottom line is percentage "certain" responses (conservative attitude). Middle line is "certain" responses plus "hunches" (liberal attitude). Top line is first two plus "guesses" (radical attitude).

gories one, two, and three provided us with a new attitude toward reporting, which we have called the "radical" attitude.

Figure 10 shows a plot of the cumulative percentage of reports (average for ten observers) under each of these three attitudes. Although these curves are taken from responses to the same signals, the lower curves (for conservative and liberal attitudes) are quite like those for groups working under only one attitude. A comparison of Fig. 10 and 3 shows that the four-category groups were slightly superior to the average of all observers working under a single attitude. Their false report rates, under both liberal and conservative attitudes, were a little lower, but for most signal strengths, the percentage of reports was a little higher. This difference is probably attributable to the fact that these two groups were particularly highly motivated.

The table presenting individual data for each of the ten observers may be obtained from the American Documentation Institute.[3] Each man made from 130 to 160 judgments at each signal strength (three times that many on blanks) so the individual curves are fairly stable. The table shows individual differences. Observers 4 and 9 were consistently low in their report rates. Observer 7 was much too free in reporting under the conservative attitude (8.1 per cent false reports), although his liberal and radical curves are comparable to the rest of the group. The remaining seven observers are fairly homogeneous, and separate averages are given for this subgroup. The average of all 10 observers, corrected for guessing, is given. A comparison of the corrected averages for the liberal and radical attitudes shows that this additional category of judgment also adds something more than sheer guessing, although the addition in this case was not as large as in the shift from conservative to liberal attitudes.

TABLE 1

RATE OF INFORMATION TRANSMISSION IN BITS PER SIGNAL AND RESPONSE RATES FOR INDIVIDUAL OBSERVERS MAKING FOUR-CATEGORY JUDGMENTS

Observer	Rate of Information Transmission in Bits/Signal			Response Rate		
	Cons.	*Lib.*	*Rad.*	*Cons.*	*Lib.*	*Rad.*
1	.103	.098	.069	.369	.646	.788
2	.081	.072	.038	.320	.588	.840
3	.095	.072	.023	.370	.603	.806
4	.065	.074	.030	.254	.440	.654
5	.104	.077	.041	.362	.617	.812
6	.110	.070	.079	.388	.543	.684
7	.088	.081	.062	.441	.557	.676
8	.097	.078	.047	.374	.582	.717
9	.061	.043	.024	.281	.497	.665
10	.091	.092	.074	.331	.562	.712
Avg.	.090	.076	.049	.349	.564	.735

[3] For table of individual data for observers making four-category judgments, order Document 3932 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting $1.25 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.25 for photoprint readable without optical aid.

Since the recording of four response categories took two of the five rows on the teletype tape, only two observers could be run simultaneously under this condition. Consequently, correlations between observers in these groups are based on too few observations to be at all reliable. They were computed, however, and were comparable to those reported in the section on intercorrelations of observers.

Information rates were also calculated for each of the ten observers under each of the three attitudes. The results are shown in Table 1. Again it may be seen that the average rate of information decreases as observers adopt a more liberal attitude toward reporting. Only two of the ten observers increased their rate of transmission in going from the conservative to the liberal attitude. One observer increased his rate in going from the liberal to the radical attitude.

IV. THE PROPOSED MODEL AND ITS APPLICATION

A. The Model

The statistical character of a sensory threshold has long been recognized, but only in terms of a distribution of sensitivity. That is, it is usually assumed that an observer either does or does not hear a signal, and that the variability in report comes from a variation in momentary sensitivity. The usual psychophysical experiment does not even test the observer to see if he will report when there is no objective signal. If the test for false reports is used, the resulting reports are treated simply as errors, to be kept to a minimum.

We have shown that the false report should be treated as something more than an error in response, since observers allowed to make false reports show a real gain in sensitivity to signals. Fig. 3 showed the average cumulative response curve, under both liberal and conservative instructions toward reporting, corrected for the guessing implied by false report rate. If nothing but guessing was involved in the false reports, these two curves should have come together. The correction assumes that the observer either does or does not hear the signal. He reports positively when he does, and also reports positively, with a certain probability, when he doesn't. This probability can be taken as equal to the false report rate, since the signal should never be heard under these conditions. The failure of this correction to bring the curves together indicates the inadequacy of these assumptions.

As an alternative to the usual two-dimensional treatment of threshold, we wish to propose a three-dimensional model. The dimensions are signal-to-noise ratio (or signal strength), subjective intensity of signal, and probability density of that subjective intensity. Figure 11 gives an oversimplified diagram of

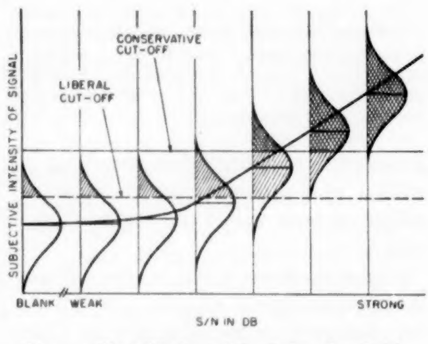

Fig. 11. Simplified model of the threshold.

the model. It shows, for each signal-to-noise ratio, a probability distribution of subjective intensity. It is assumed in the model that such a distribution exists for each signal-to-noise ratio, including blanks, that each distribution is normal, and that all have the same variance. The means of the distributions are here assumed to decrease linearly with S/N in db, but to decrease nonlinearly below

some point. When the signals are very weak with respect to the noise, it matters little whether there is actually a signal or not. The blank (no signal) is assumed to be just the limiting case of weak signals.

The effect of attitude toward reporting is then to move a cut-off line along the dimension of subjective intensity. Any signal falling above this line will be reported, so that the cross-hatched areas represent the percentage of reports under conservative instructions, and the ruled areas are the corresponding percentages for liberal instructions.

Although these assumptions are not quite adequate, it is instructive to see how they compare with the usual assumptions. If we assume an all-or-none threshold, normally distributed with respect to signal-to-noise ratio in decibels, this model would be modified to the extent of straightening out the line relating the means of the normal distributions to S/N. Since the case of no signal represents a negatively infinite signal-to-noise ratio, this means that there should be no false reports. We found it impossible to prevent false reports completely although we could, by instruction, manipulate the rate at which they occurred. Moreover, all our plots of cumulative percentage of signals reported (even those "corrected for guessing") showed a deviation from normality in the lower end of the scale. It could be argued that this result could be more simply justified by changing the assumption about the physical scale against which we should expect normality of threshold distribution. If we used some transformation of S/N in db, we could normalize the lower branch of the distribution. However, no monotonic transformation will bring the blanks into line.

The main objection to this proposed model stems from its extreme flexibility. By changing the assumptions about the function relating the means of the distribution to S/N, or the assumptions about the relative variances of the distributions, or the assumptions about normality of distribution, it is possible to fit the model to almost anything. However, the two-dimensional model does not account adequately for our results, and if we can find a consistent set of assumptions for our model, it should prove useful.

1. Assumptions on Variability

Our first concern is with the variances of the distributions at different signal strengths. In the model as it is shown in Fig. 11, they are assumed to be constant. This assumption turns out to be inadequate, and needs to be investigated. It is convenient to divide the variance into two components, the variability common to observers and the individual variability.

a. Variance Common to All Observers

It was shown in the results section that observers listening to the same signal and the same noise tend to report at the same time and to fail to report at the same time. This correlation may be due in part to factors such as shifts in group morale, fatigue, and sequence of signals, but the principal source is the noise. Since all five observers are listening simultaneously, any random fluctuation in the noise will affect all of them to the same extent. The relative magnitude of this common variance is indicated by the size of the correlation between observers.

If we assume that subjective intensity is a random variable composed of the sum of noise intensity and individual variation, and also that individual variation

is the same for two observers, then the correlation coefficient, r, will be the percentage of the total variance that is attributable to noise. The noise is always the same, so this is one feature of the situation that is independent of signal strength or of attitude. In the case of observers working under the conservative attitude, it is assumed that the value of r remains constant at .55 regardless of signal strength. It follows, then, that the total variance of the distribution of subjective intensity is independent of signal strength in this case.

When observers are working under the liberal attitude, however, the correlation between observers is a function of signal strength, being lower for weak signals than for strong signals. This means that the noise variance, which is independent of the effects of attitude, is now a smaller percentage of the total variance. This, in turn, shows that the other factor contributing to the total variance, the individual variance, is greater for weak signals than for strong.

b. Individual Variability

The sources of individual variability are not quite so easy to identify. There should be at least two, although finding objective evidence for either is quite difficult. There is variation in sensitivity and variation in cut-off level.

There is no direct evidence for variation in sensitivity, but it seems to be a theoretical possibility. Actually, it is probably not sensitivity to an 800-cycle tone that is varying, since such a variation should be accompanied by variation in sensitivity to the masking noise in the critical band. If this were the case, signal-to-noise level should remain unchanged, and the probability of a report should not change. However, it is reasonable to assume that the receptor mechanism itself contributes varying amounts of internal noise, which could change the signal-to-noise level.

In terms of our model, such variation in internal noise could contribute to the 45 per cent of the total variance that is associated with the individual working under conservative instructions. This part of the individual variance could not change with instructions, since the observer can make judgments with respect to both attitudes simultaneously.

The second source of individual variability, variation in cut-off level, is the one that must vary with instructions. The intercorrelations of observers show that individual variance on weak signals is greater under liberal than under conservative instructions. The external noise is independent of attitude, and so is internal noise, if it exists. This process of elimination leaves cut-off variability as the only source of a difference.

Variation in cut-off level has an effect that is easy to handle. The result of variation in cut-off level is mathematically the same as that of a stable cut-off level on a distribution whose variance is the sum of the variance of the original distribution and the variance of the cut-off distribution.

Given two normal distributions of x and y (representing here the distribution of subjective intensity and the distribution of cut-off level, respectively), with variances σ_x^2 and σ_y^2, we want to find the probability of x exceeding y. This probability will vary, of course, with the value of c. (c is the mean value of the cut-off distribution, expressed in units equal to th σ of the main distribution.)

We want $P(x > y) = P[(x - y) > 0]$. The distribution of $(x - y)$ will be normal (since x and y are both normal); with the mean of $M_x - M_y = -C$, and variance of $\sigma_x^2 + \sigma_y^2$ (x and y assumed to

be independent). Thus, if we let $u = (x - y)$,

$$P[(x-y)>0] = \int_0^\infty \frac{1}{\sqrt{2\pi}\,\sigma_u} e^{-\frac{1}{2}\left(\frac{u+c}{\sigma_u}\right)^2} du$$

Let

$$u' = \frac{u+c}{\sigma_u},$$

and

$$P[(x-y)>0] = \int_{c/\sigma_u}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{u'^2}{2}} du'$$

$$= \int_{c/\sqrt{\sigma_x^2+\sigma_y^2}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{u'^2}{2}} du$$

Thus the variable cut-off has the same effect as a constant cut-off equal to the mean of the cut-off distribution operating on a distribution of increased variance.

It would be very helpful if we could develop a theory of the change in cut-off variance as a function of signal strength. In this case we can deduce from the intercorrelations that it must change, and can even estimate how much it changes. However, in the case of an absolute threshold there could be little common variance, and hence no way to estimate changes in variance with changes in attitude or signal strength.

Since the observer does not know the actual signal strength at the time he makes his judgment, it is a little unreasonable that his variability is related to signal strength. We might assume that a change in attitude brings an increase in cut-off variability that is independent of actual signal strength, but which is applied to only that proportion of the judgments falling below the conservative cut-off level. This assumption would mean that most of the judgments on weak signals would be subject to the liberal variance, and most of those on strong signals to the conservative variance, so the trend in liberal intercorrelations as a function of signal strength would be in the right direction. However, the correlations that are estimated from these assumptions do not correspond very well with the observed values of r. Moreover, the resulting distribution is not normal, but consists of elements of two different normal distributions.

Since we are unable to justify anything more elaborate, we have made our assumptions as simple as possible. In terms of the model, the subjective intensity distributions under conservative instructions are assumed to have the same variance. This variance is arbitrarily taken as unity, with .55 associated with noise and .45 with individuals. For the liberal distributions, the variance is assumed to vary with signal strength. The size of the variance can be estimated from knowledge of the intercorrelations of observers. Since r is the percentage of common variance, and since the noise variance is .55, regardless of attitude,

$$r_{(\text{liberal})} = \frac{.55}{\sigma^2_{(\text{liberal})}}; \quad \text{or,} \quad \sigma_{(\text{liberal})} = \sqrt{\frac{.55}{r_{(\text{liberal})}}}.$$

2. *Other Assumptions*

One other assumption calls for reconsideration. It deals with the normality of distribution of the subjective intensity distributions. We have no real justification for this choice, and can suggest no empirical check. However, it seems to provide a good approximation, and we need to make only a minor change in the assumption.

Observers working over long testing sessions find it extremely difficult to keep constantly alert. It was pointed out

above that in a few instances in this experiment observers were known to have fallen asleep for short periods. There were undoubtedly other instances, unknown to the experimenter, in which the observer lapsed into a state that might well have been sleep, even if his eyes were open. Better testing conditions and more careful selection of observers might have improved this situation, but it is a factor in any threshold experiment.

The effect of this lapse is to put the observer into another state, as far as probability of report is concerned. He may occasionally awake and, realizing that he has missed a signal, guess that one occurred. This might happen occasionally under the radical attitude, but more rarely under the liberal, and not at all under the conservative attitude. We assume, therefore, that the subjective intensity distribution is actually bimodal, with the larger part of the area (98 per cent or more) lying under a normal shaped curve. The remaining 2 per cent or less of the area falls far below, out of reach of any cut-off line. When most of the area lies below the cut-off line, this assumption will have negligible effect. However, for strong signals, the effect is more important.

B. Application of the Model

1. *Application to Average Individual Curves*

We have spoken throughout about the distribution of subjective intensity. This is a convenient fiction, designed to facilitate discussion. Our measurements are of percentage of report, and we assume only that there is some normally distributed variable underlying the report, and that the integral of this distribution above the cut-off level gives the percentage of reports.

To fit the model to the data we have fixed the means of the distributions with respect to a constant conservative cut-off level. We assume that all the conservative distributions have the same variance, and since we have no independent scale on which to measure this variance, we arbitrarily take it as unity. This allows us to enter the table of the integral of the normal curve, find the value of the integral equal to our per cent of report for a given signal, and read off the corresponding x/σ score. The mean of that distribution must be this number of units above (or below) the conservative cut-off line. The resulting plot of means has the same form as a plot of the percentages on normal probability paper.

Once the distribution means have been fixed, the liberal cut-off line can be plotted. Because of the assumed variability of the liberal cut-off line, the σ of the liberal distributions is no longer unity, but varies with signal strength. The value of the σ is obtained by computing $\sigma_L = \sqrt{.55/r}$, where r is the correlation between observers for a given signal strength. Rather than use the observed values of r, we drew a freehand smoothed curve through these values, and used the values from the smoothed curve. The observed points were quite variable, and the smoothed curve cannot be regarded with a great deal of confidence. It followed roughly the form of an ogive.

After the σ of the liberal distribution has been obtained, the distance from mean to liberal cut-off line can be gotten by looking up the x/σ value for the observed percentage of liberal reports, and multiplying by σ. If the model fits, all of these cut-off points should fall on a horizontal line.

One further correction is needed. Because of the occasional lapses of ob-

TABLE 2

COMPUTATIONS FOR FITTING THE MODEL TO DATA FROM THE MAIN EXPERIMENT

S/N	Conservative			Liberal			
	%Rep.	*Corr.*	*Z*	*% Rep.*	*Corr.*	*r*	*Z*
Blank	.035	.036	+1.80	.259	.264	.25	+ .93
9.5	.059	.060	+1.56	.306	.312	.34	+ .62
10.5	.055	.056	+1.59	.336	.342	.35	+ .51
11.5	.072	.073	+1.45	.361	.368	.36	+ .42
12.5	.088	.090	+1.34	.370	.377	.37	+ .38
13.5	.111	.113	+1.21	.438	.446	.38	+ .17
14.5	.145	.148	+1.05	.481	.490	.39	+ .04
15.5	.199	.203	+ .83	.538	.548	.41	− .14
16.5	.276	.281	+ .58	.609	.620	.42	− .35
17.5	.368	.375	+ .32	.699	.712	.44	− .63
18.5	.496	.505	− .01	.779	.793	.46	− .89
19.5	.659	.671	− .44	.868	.884	.47	−1.30
20.5	.806	.821	−.92	.928	.945	.48	−1.71
21.5	.897	.913	−1.36	.961	.978	.49	−2.13
22.5	.945	.962	−1.77	.972	.9898	.50	−2.44
23.5	.971	.989	−2.29	.981	.99895	.50	−3.22

Note: *r* taken as .55 for conservative observers throughout.
The correction is for an assumed 1.8% response failures.

servers, it is assumed that the distribution is actually bimodal, with the main portion of the curve being normal in form, and a small part of the area way down on the scale, where no cut-off line could reach it. We want, therefore, to deal only with the main, upper curve, and want to adjust its area to unity. If 1.8 per cent of the area is assumed to fall in the smaller part of the curve, for example, then all areas should be multiplied by the reciprocal of .982 to give a corrected report.

The computations for the model-fitting to the data from the main experiment are given in Table 2. The correction factor applied to the report rates was 1/.982 or 1.0183. This assumes that 1.8 per cent of the judgments fell outside the distribution, which is probably too high a figure. This choice was arbitrary and was made to make the report of the strongest signal under the liberal instructions nearly unity. The corrected plot is shown in Fig. 12 and it may be seen that the liberal cut-off points fall close to a straight line, one sigma below the conservative cut-off. There is a slight tendency for the cut-off points to rise for the strong signals, despite the correction. The uncorrected data provide a plot that is, on the scale of this graph, indistinguishable from the corrected one below 18.5 db. Above this point, the uncorrected graph shows the means of the distributions sagging down toward the conservative line. For example, the dis-

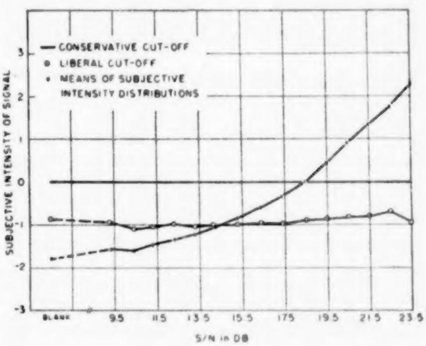

FIG. 12. Model applied to average data, all observers. For conservative observers, *r* taken as .55, independent of signal strength. For liberal observers *r* taken from smooth curve fitted to Fig. 6.

tribution mean for the strongest signal is is 2.29 corrected, and 2.90 uncorrected. The corrected liberal cut-off point is −3.22; uncorrected it is −2.17, or only .27 below the conservative cut-off, instead of the 1.00 that held for the weaker and middle-range signals.

2. *Application to Four-Category Judgments*

The same sorts of curves were plotted for the ten observers who used the four-category judgment. In this case, the area in the smaller, excluded distribution was taken as .006, instead of .018. This figure was chosen to make the percentage of report of the strongest signal under the radical attitude nearly one. The uncorrected percentage was .9935, and the corrected percentage .9995. Obviously, the percentages were not observed with this degree of accuracy, so that the correction is fairly arbitrary.

The correlations used were taken from the main experiment for the liberal curve. For the radical attitude, there were no correlational data. The correlations used were taken from the smoothed curve for liberal correlations, with the transformation $r_R = 2r_L - .55$. This transformation preserves the form of the curve

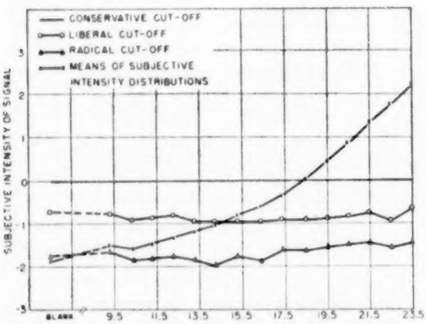

FIG. 13. Model applied to data from 10 observers making four-category judgments.

relating liberal correlations to signal strength, and doubles its distance from the conservative correlation.

The resulting curves are plotted in Fig. 13. Again, the fit is fairly good for the liberal cut-off line, although there tends to be a rise in the line for the stronger signals. The radical cut-off line shows this effect to a much more marked extent. There are several possible reasons for this rather poor fit on the radical cut-off. Our estimates of the variances of the radical cut-offs were complete guesswork. We may have underestimated the variances of the distributions for stronger signals. It is also possible that we have underestimated these variances in the case of the liberal judgments as well. Those variances were estimated from the intercorrelations of observers, but as the signals got stronger, more of the correlations were indeterminant. This indicates the possibility of a systematic bias, since only the observers who failed to respond got into the correlations. This situation could be brought about by occasional signal failures, or by the fact that some observers were less sensitive to the signals and were, in effect, listening to signals a few db weaker than those presented to other observers.

Another possible reason for the failure of the fit of the radical cut-off line lies in our assumption of normality of the distributions of subjective intensity. This assumption fits fairly well in the middle range of the distribution, but it is quite possible that the assumption fails at the extremes. The actual form of the distribution need not deviate greatly from normality. It should be kept in mind that this method of presenting the data greatly magnifies deviations at the extreme of the distribution. In terms of deviation of observed percentage report from the expected value, the fit is not

bad. For example, in Fig. 13 the average percentage report was .9935 instead of .9995 for the strongest signal under the radical attitude. If there had been only one of the 1550 signals missed, instead of 11, the radical cut-off would have fitted without correction. However, the systematic nature of the deviation as plotted would seem to indicate some faulty assumption.

3. Application to the Concept of Information Transmission

It was shown in the section on results that the conservative observers transmitted more information. There was some indication, however, that this statement is not universally true. The extreme of conservatism would be the reporting of no signals or blanks, whereas the extreme of liberalism would be a report at every opportunity—all signals and all blanks. Neither of these extreme modes of response would transmit any information, so that, logically, we must have at least one intermediate mode of response that transmits a maximum amount of information.

Application of the model proposed here makes it possible to plot rate of information transmission as a function of false report rate. It is assumed that as the cut-off lines are moved down, the variances of the individual distributions of subjective intensity increase in a systematic fashion. The values of these variances were estimated from interpolation or extrapolation of the observed values. For example, for any false report rate less than 3.5 per cent the intercorrelations between observers were assumed to be .6 for all signal strengths. For false report rates greater than 3.5 per cent the correlations were assumed to decrease toward the values obtained for the liberal observers. The expected report percentage of all signals was then computed for each false report rate and the information transmission rate was computed.

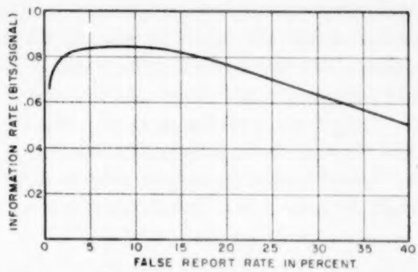

FIG. 14. Calculated (theoretical) values of information transmission plotted against false report rate.

The result is shown in Fig. 14. The graph shows that a single maximum exists, somewhere near 5 per cent false reports. Any false report rate less than 2 per cent leads to some loss of information, and suppression of false reports much below 1 per cent leads to serious loss.

4. Application to Multiple Observers

If each of the five individuals has a probability p of reporting a signal of a given strength, then the probability that at least one will report it is given by one minus the probability that all will fail to report. The joint probability that all will fail to report when the cut-off line is at a given value a, is the integral of the normal correlational surface up to the value a on each of the five variables. The probability that at least two (two or more) will report is this first probability minus the probability that four will be below the cut-off and one above. This latter probability is given by another segment of volume under the normal correlational surface. Similarly, the rest of the n-or-more curves may be obtained.

The method of integration of the nor-

mal correlational surface discussed in the introduction allows us to get all these values. The general procedure was that of summing, for all values of t in steps of .1, the product of the normal ordinate and the area to the fifth power from $t - a$ to infinity, for a normal distribution with $M = 0$, $\sigma = 1$. Schematically,

$$(\text{Vol. up to } a) = \frac{1}{\sigma_6} \sum_{t=-\infty}^{\infty} \left(\text{ord. at } \frac{t}{\sigma_6}\right) (\text{Area from } t-a \text{ to } \infty)^5.$$

By taking this sum for lower powers of the area, all the marginals of the five-variable surface can be computed. From the marginals it is possible to build up the volumes in the five dimensional surface for all possible splits. By changing the value of σ_6 ($\sigma_6 = \sqrt{r/\ (1\text{-}r)}$) it is possible to get the integral for any value of r. These sums must be computed for enough values of a to plot the cumulative distribution of the n-or-more criteria. For the case of the conservative observers this is relatively simple, since the correlation is the same for all values of a (r is independent of signal strength). For the liberal observers it was necessary to use different values of r for each value of a.

In the model we have proposed, the distributions of subjective intensity refer to individual performance. To get the percentage of times n-or-more individuals respond to a given signal we replace the individual curves with the proper group-distribution. For one-or-more observers out of five, for example, the new distributions will have means that are shifted upward, and will be skewed, with the tail running downward. The curve for all five will be symmetrical to the one-or-more, with mean shifted down, and tail running upward. The two-or-more and four-or-more will also be symmetrical to each other, but shifted less. The three-or-more distribution will have the same mean as the individual distributions, but will have a smaller variance. Thus, the criterion of three-or-more reporting should yield 50 per cent reports for the same signal strength that gives 50 per cent for the individual. However, the group should report fewer of the weak signals, and more of the stronger ones. Examination of the data shows this to be the case. The 50 per cent points for three-or-more and for individuals correspond as closely as the graphs can be read for both the conservative and the liberal observers.

Before discussing the curves so constructed, it is necessary to consider the effect of the correction for inattention we made in applying the model to individual performances. This correction assumed that the observers were effectively out of the system for a small percentage of the judgments (1.8 per cent in the main experiment). This left only .982 as the area under the main distribution, and the observed percentages were divided by .982 to make the area unity. This correction could also be considered as an adjustment for a slight non-normality of the distribution.

The effect of this assumption on the n-or-more curve is similar to its effect on individual observers. In the case of two observers, Fig. 15 illustrates the result. If the inattentiveness of the two observers is assumed to be independent, then the joint distribution will be as pictured, and the main surface will have a volume of $1 - 2 \times .018 = .964$ under it. The volume under this surface from a to ∞ on both axes is the computed value for the normal surface multiplied

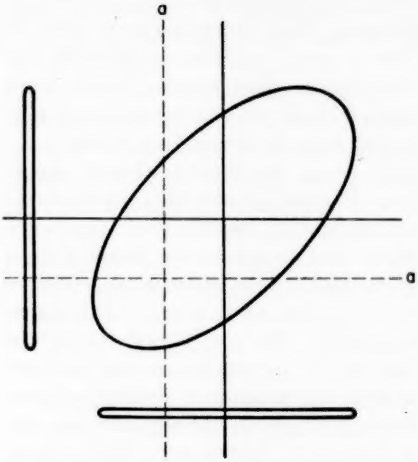

FIG. 15. Assumed effect of inattentiveness on joint distribution of two observers.

by .964. Similarly, for three variables, the correction consists of multiplying by $1 - 3 \times .018 = .946$ and so on. In other words, if each observer is in a state of nonresponsiveness part of the time, the number of times we can expect all observers to respond to the same signal will be decreased. If the individual never responds more than 98.2 per cent, then the group of five will never all respond to the signal more than 91.0 per cent. The corrected marginals allow computation of the cumulative distributions for all possible combinations of response.

This correction has very little effect except in the case of the curve for all-five-responding. This one is uniformly 91 per cent of its uncorrected height. However, this does not quite complete the account of the correction. These cumulative curves, corrected and uncorrected, are laid out on a scale with zero at the mean of the individual distribution, and units equal to the standard deviation of the individual. To get the expected uncorrected percentage report, we used the individual percentage report, converted this to an x/σ score, and read off the corresponding values from the uncorrected cumulative *n*-or-more curves. For the corrected percentage report for *n*-or-more observers, we used the corrected individual percentages (from Table 2). Thus the correction works in two ways, and the net effect is to increase very slightly the expected percentage report for all the curves except that for all-five-reporting. This one is lowered a little (3 per cent at the strongest signal) by the correction.

Figures 16 and 17 show, for the conservative and liberal observers, respectively, the corrected theoretical curves, with the observed values plotted on them. Because of the variation of r with signal strength in the case of the liberal observers, computation of the theoretical values was quite laborious and, in general, only

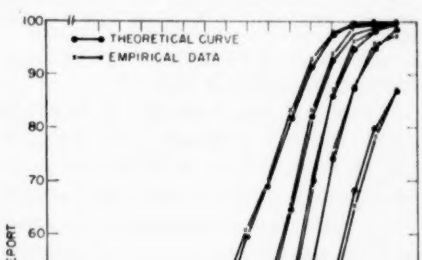

FIG. 16. Percentage report by *n*-or-more observers, conservative attitude. Solid points are average data from main experiment. Predicted values discussed in text.

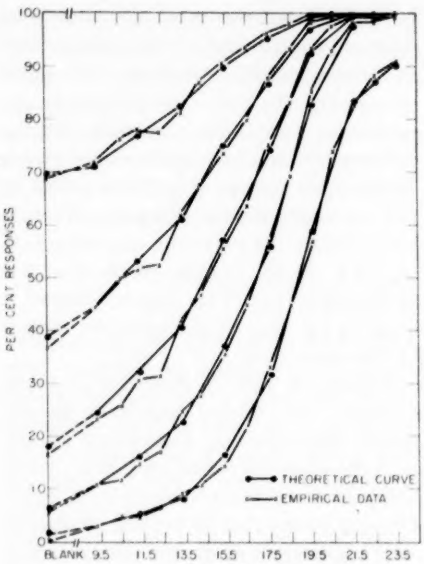

FIG. 17. Percentage report by n-or-more observers, liberal attitude. Data and predictions as in Fig. 16.

every other point is plotted.

There is a very high degree of agreement between expected and observed values. Some degree of agreement is to be expected, since the theoretical curves are based on the performance of the same individuals that made up the groups. However, we feel that the agreement helps to verify the general approach.

No new light is thrown on the validity of our model. Since the individual percentage report was used to predict the behavior of the group, we have used the cut-off lines shown in Fig. 12. However, had we straightened out the liberal cut-off line and used it, the difference in expected group performance would be changed so little as to be indistinguishable on the scale of our graphs.

5. *Extensions*

In terms of practical usefulness, the ideas explored here have been disappointing. Our results indicate that the liberal and conservative attitudes are very nearly equal in effectiveness when group criteria are selected to equate false report rates. It should be possible to deduce from the theory whether this is generally true, or only happens to be true for the special conditions of this experiment. Unfortunately, the extreme labor of computing the theoretical n-or-more curves makes such a deduction almost impossible. We can speculate, on the basis of the calculations of rate of information as a function of false report rate, that an extremely conservative attitude is inefficient, but over a fairly broad range of false report rates, attitude makes little difference.

The gain of the group over the individual is also disappointingly small. It is true that the conditions of this experiment were such as to minimize the gain of the group over the individual, but it is doubtful that conditions which maximize the gain will greatly increase it. The conditions for increasing the gain are those which increase individual variance, or which reduce correlation between individuals. That these conditions do not always produce a gain was shown in this experiment. Shifting from the conservative to the liberal attitude met both conditions, but did not increase the superiority of the group over the individual.

For five conservative observers, the criterion of two-or-more reporting will always give the best approximation to the individual false report rate, hence we can confine our attention to it. If the correlation between observers could go to zero while everything else remained unchanged, there would be a gain of roughly 1.5 times what it is now for the stronger signals. Referring to Fig. 5 we can see that this means about 1.5 db at the 50

per cent level, or 3 db at the 95 per cent level. Below the 25 per cent level there is actually a slight loss.

The factors which would tend to reduce correlation (slower rate of presentation of signals, variation in signal characteristics, etc.) should also tend to increase individual variability, and it would seem that the gain should be even greater. However, the interaction of the two factors is such that the gain does not increase.

Suppose that signals were presented less often. We would expect to find variability increased, presumably because of an increase in individual variability (variation in cut-off level). On the scale of our model (see Fig. 12) the total variance for conservative observers is one, with σ_c^2 (the common variance) $= .55$, and σ_I^2 (individual variance) $= .45$ for conservative observers. Suppose that the individual variances were quadrupled. It is difficult to make a realistic set of assumptions about this change in the cut-off variability that will leave it normally distributed, but we will assume that all cut-off scores are subject to the transformation $I' = 2I + 1.33$. This will increase the variance by 4 and increase the mean cut-off by 1.33. The mean increase is arbitrary, chosen to give 2 per cent false reports, since anything that increases the variability this much might be expected to reduce the over-all report rate. Our model would now have the same plot of mean subjective intensity, but the variance of each distribution would be $\sigma_c^2 + \sigma_I^2 = .55 + 1.80 = 2.35$ and $\sigma = 1.53$. The new cut-off line would be 3.14 units above the mean for the blanks. This yields an x/σ score of 2.05, and 2 per cent false reports. The expected percentage reports can then be obtained by reading from the graph (Fig. 12) the distance from the mean to the new cut-off line, and dividing each reading by 1.53 to get the new x/σ score. Entering the table of area under the normal distribution, we can get the expected percentage report. These figures, plotted against signal strength, will give a curve that rises less steeply than the present one, reaching only 72 per cent report for the strongest signal.

The correlation between observers under the assumed conditions would be $r = \sigma_c^2/(\sigma_c^2 + \sigma_I^2) = .55/2.35 = .23$. We can then estimate the expected percentage report for two-or-more observers, out of a group of five. The resulting curve shows a gain of about 1.5 db at the 50 per cent level, and 3.5 db at the 95 per cent level. Again the gain of the group over the individual is about 1.5 times the gain shown in our experimental results.

Similarly, we can estimate what should happen if individual variability remains as it was in our experiment, but all sources of common variance drop out. This, again, represents an extreme, rather than a reasonable, assumption. For an assumed false report rate of 3.5 per cent, the plot of expected report against signal strength now rises more steeply than in our experiment, reaching 95 per cent at about 18 db on the scale of our graphs. Here the steep slope of the curve leaves little room for improvement, and the expected gain of the group over the individual is approximately the same as it was in our results.

These calculations are rough, and are based on several very dubious assumptions, but they indicate that no gain of great practical significance can come from the use of a team of observers.

V. SUMMARY

This paper aimed at the presentation of an adequate theory of the multiple observer; that is, of the effectiveness of a group of observers working as a team. Before this could be done, it was necessary to develop a model of the threshold that takes false reports (i.e., positive reports when no signal is presented) into consideration.

The model proposed is three-dimensional, with a probability distribution of subjective intensities for each signal strength, including zero strength. Variation of the observer's attitude toward reporting moves a cut-off line up or down the dimension of subjective intensity. Any signal (or blank) exceeding this cut-off is reported. The necessity for such a model is indicated by conclusive evidence that the "guesses" made under the more liberal attitudes cannot be considered to be sheer guesses. If they were, a correction for guessing based on the false report rate should superimpose cumulative report curves, regardless of attitude. The data consistently failed, by a wide margin, to meet this assumption.

Two sets of experimental data were fitted to the model. In the main experiment 23 groups of five observers each listened for an 800-cycle tone of .41-sec. duration, against a background of broadband noise. Twelve of these groups worked under instructions to be very conservative in reporting, and 11 groups were given more liberal instructions. The conservative groups were encouraged to avoid all false reports, but made an average of 3.5 per cent. The liberal groups were encouraged to guess whether or not they heard a signal, and had a false report rate of 25.9 per cent.

In a second experiment, ten observers were each given a large number of trials under instructions to make a four-category judgment on each tone. These observers made one response if they were sure they heard a tone, another if they thought they heard it, and for the remaining signals they guessed tone or blank.

The threshold model was applied by deducing variances of subjective intensity distributions from the intercorrelations of observers in the main experiment. The conservative cut-off line was taken as a reference point, and the cut-off points for other attitudes were plotted with respect to it. In both experiments the data fit the model fairly well, although there was some evidence of a discrepancy for the stronger signals.

With this model as a basis, a theory of the multiple observer was worked out and applied to the data from the main experiment. Agreement of theory and observation was satisfactory. Prospects for a practical application of the multiple observation unit were less satisfactory. Under the conditions of this experiment, the gain of the group over the individual was slight when false report rates were equated. The gain cannot be expressed adequately in terms of a single figure, since it was greater for stronger signals. However, at the 95 per cent report level, the group was only two db better than the individual.

REFERENCES

1. HASTINGS, C., MOSTELLER, F., TUKEY, J. W., & WINSOR, C. P. Low moments for small samples: a comparative study of order statistics. *Ann. math. Stat.*, 1947, **18**, 413-426.
2. MOOD, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
3. NATIONAL DEFENSE RESEARCH COUNCIL, DIV. 17, SEC. 3. Response characteristics of interphone equipment. O.S.R.D. Report No. 3105, 1944.
4. PEARSON, K. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philos. Trans. roy. Soc.*, Ser. A, 195, 1901, 1-47.
5. PEARSON, K. *Tables for statisticians and biometricians.* Part II. London: Cambridge Univer. Press, 1931.
6. SCHAFER, T. H. Detection of a signal by several observers. Report No. 101, U. S. Naval Electronics Laboratory, San Diego, Calif., 1949.
7. SCHAFER, T. H. Influence of the preceding item in measurements of the noise-masked threshold by a modified constant method. *J. exp. Psychol.*, 1950, **40**, 365-371.
8. SHANNON, C., & WEAVER, W. *The mathematical theory of communication.* Urbana: Univer. of Ill. Press, 1949.

(Accepted for publication February 16, 1953)